

# Source

G.S. Boolos, J.P. Burgess, R.C Jeffrey:  
Computability and Logic. Cambridge University Press 2002.

# Theories

A **signature** is a (finite or infinite) set of predicate and function symbols. We fix a signature  $S$ . “Formula” means now “formula over the signature  $S$ ”.

A **theory** is a set of formulas  $T$  closed under consequence, i.e., if  $F_1, \dots, F_n \in T$  and  $\{F_1, \dots, F_n\} \models G$  then  $G \in T$ .

**Fact:** Let  $\mathcal{A}$  be a structure suitable for  $S$ . The set  $F$  of formulas such that  $\mathcal{A}(F) = 1$  is a theory.

We call them **model-based** theories.

**Fact:** Let  $\mathcal{F}$  be a set of closed formulas. The set  $F$  of formulas such that  $\mathcal{F} \models F$  is a theory.

# The signature of arithmetic

The signature  $S_A$  of arithmetic contains:

- a constant  $0$ ,
- a unary function symbol  $s$ ,
- two binary function symbols  $+$  and  $\cdot$ , and
- a binary predicate symbol  $<$ .

(slight change over previous definitions)

**Arith** is the theory containing the set of closed formulas over  $S_A$  that are true in the canonical structure.

**Arith** contains “all the theorems of calculus”.

# Decidability, consistency, completeness, ...

A set  $\mathcal{F}$  of formulas is **decidable** if there is an algorithm that decides for every formula  $F$  whether  $F \in \mathcal{F}$  holds.

Let  $T$  be a theory.

$T$  is **decidable** if it is decidable as a set of formulas.

$T$  is **consistent** if for every closed formula  $F$  either  $F \notin T$  or  $\neg F \notin T$ .

$T$  is **complete** if for every closed formula  $F$  either  $F \in T$  or  $\neg F \in T$ .

$T$  is **(finitely) axiomatizable** if there is a (finite) decidable set  $\mathcal{X} \subseteq T$  of **axioms** such that **every** closed formula of  $T$  is a consequence of  $\mathcal{X}$ .

# Conventions and notation

In the following: set of axioms = decidable set of formulas over  $S_A$

$T_{\mathcal{X}}$  denotes the theory of all closed formulas that are consequences of a set  $\mathcal{X}$  of axioms.

# Basic facts

**Fact:** Every theory contains all valid formulas (because they are consequences of the empty set).

**Fact:** Model-based theories (like `Arith`) are consistent and complete.

**Fact:**  $T$  is consistent iff there is a closed formula  $F$  such that  $F \notin T$ .

**Proof:** If  $T$  is consistent then  $F \notin T$  for some closed  $F$  by definition. If  $T$  is inconsistent, then there exists a formula  $F$  such that  $F \in T$  and  $\neg F \in T$ . Let  $G$  be an arbitrary closed formula. Since  $F, \neg F \models G$  and  $T$  is closed under consequence, we have  $G \in T$ .

# Basic facts

**Lemma:** If  $T$  is axiomatizable and complete, then  $T$  is decidable.

**Proof:** If  $T$  inconsistent then  $T$  contains all closed formulas, and the algorithm that answers “ $F \in T$ ” for every input  $F$  decides  $T$ .

If  $T$  consistent, let consider the following algorithm:

- Input:  $F$

Enumerate all syntactic consequences of the axioms of  $T$ , and for each new syntactic consequence  $G$  do:

- If  $G = F$  halt with “ $F \in T$ ”
- If  $G = \neg F$  halt with “ $F \notin T$ ”

Observe: the syntactic consequences of the axioms can be enumerated.

We prove this algorithm is correct:

- If algorithm answers “ $F \in T$ ”, then  $F \in T$ .  
If algorithm answers “ $F \in T$ ”, then  $F$  is syntactic consequence, and so consequence of the axioms. Since  $T$  is a theory,  $F \in T$ .
- If algorithm answers “ $F \notin T$ ”, then  $F \notin T$ .  
If algorithm answers “ $F \notin T$ ”, then  $\neg F$  is consequence of the axioms and so  $\neg F \in T$ . By consistency,  $F \notin T$ .
- The algorithm terminates.  
Since  $T$  is complete, either  $F \in T$  or  $\neg F \in T$ .  
Assume w.l.o.g.  $F \in T$ .  
Since  $T$  is axiomatizable,  $F$  is a consequence of the axioms.  
So  $F$  is a syntactic consequence of the axioms.  
So eventually  $G := F$  and the algorithm terminates.



# Basic facts

**Theorem:** Arith is **undecidable**.

**Proof:** By reduction from the halting problem, similar to undecidability proof for validity of predicate logic.

**Theorem:** Arith is **not axiomatizable**.

**Proof:** Since Arith is undecidable, consistent, and complete, it is not axiomatizable (see Lemma).

# Gödel's first incompleteness theorem

**Theorem:** Let  $\mathcal{X}$  be any set of axioms such that  $\mathcal{X} \subseteq \text{Arith}$ .  
Then the theory  $T_{\mathcal{X}}$  is **incomplete**.

**Proof:** Since  $\text{Arith}$  is not axiomatizable, there is a formula  $F \in \text{Arith}$  such that  $\mathcal{X} \not\models F$  and so  $F \notin T_{\mathcal{X}}$ .

Assume now  $\neg F \in T_{\mathcal{X}}$ . Then  $\mathcal{X} \models \neg F$  and since  $\mathcal{X} \subseteq \text{Arith}$  we get  $\neg F \in \text{Arith}$ , contradicting  $F \in \text{Arith}$ .

So  $F \notin T_{\mathcal{X}}$  and  $\neg F \notin T_{\mathcal{X}}$ , which proves that  $T_{\mathcal{X}}$  is incomplete.

# Gödel's first incompleteness theorem

Observe:  $F \in \text{Arith}$ , i.e.,  $F$  is true in the canonical structure, but its truth cannot be proved using any set  $\mathcal{X}$  of axioms (unless some axiom is itself not true!)

In other words: for every set of true axioms, there are true formulas that cannot be deduced from the axioms

But we have no idea how such formulas look like ...

**Goal:** given a set of axioms  $\mathcal{X} \subseteq \text{Arith}$ , construct a formula  $F \in \text{Arith}$  such that  $F \notin T_{\mathcal{X}}$

# Minimal arithmetic

Minimal arithmetic  $\mathcal{Q}$  is the axiom-based theory over  $S_A$  having the following axioms:

$$(Q1) \quad \forall x \quad \neg(0 = s(x))$$

$$(Q2) \quad \forall x \forall y \quad s(x) = s(y) \rightarrow x = y$$

$$(Q3) \quad \forall x \quad x + 0 = x$$

$$(Q4) \quad \forall x \forall y \quad x + s(y) = s(x + y)$$

$$(Q5) \quad \forall x \quad x \cdot 0 = 0$$

$$(Q6) \quad \forall x \forall y \quad x \cdot s(y) = (x \cdot y) + x$$

$$(Q7) \quad \forall x \quad \neg(x < 0)$$

$$(Q8) \quad \forall x \forall y \quad x < s(y) \leftrightarrow (x < y \vee x = y)$$

$$(Q9) \quad \forall x \forall y \quad x < y \vee x = y \vee y < x$$

# Peano arithmetic

Peano arithmetic  $\mathcal{P}$  is the axiom-based theory over  $\mathcal{S}_A$  having Q1-Q9 as axioms plus all closed formulas of the form

$$\begin{aligned} \text{(I)} \quad \forall \mathbf{y} \quad F(0, \mathbf{y}) \wedge \forall x ( F(x, \mathbf{y}) \rightarrow F(s(x), \mathbf{y}) ) \\ \rightarrow \\ \forall x F(x, \mathbf{y}) \end{aligned}$$

where  $\mathbf{y} = (y_1, \dots, y_n)$ .

Observe: I is an axiom **scheme**; the set of axioms of  $\mathcal{P}$  is infinite but decidable.

# Some theorems of Q (and P)

$\neg(0 = s^n(0))$  for every  $n \geq 1$

$\neg(s^n(0) = s^m(0))$  for every  $n, m \geq 1, n \neq m$

$\forall x \ x < 1 \leftrightarrow x = 0$

$\forall x \ x < s^{n+1}(0) \leftrightarrow (x = 0 \vee x = s(0) \vee \dots \vee x = s^n(0))$

$s^n(0) + s^m(0) = s^l(0)$  for every  $n, m, l \geq 1$  such that  $n + m = l$

$s^n(0) \cdot s^m(0) = s^l(0)$  for every  $n, m, l \geq 1$  such that  $n \cdot m = l$

# Gödel encodings

A **Gödel encoding** is an injective function that maps every formula over  $S_A$  to a natural number called its **Gödel number**.

Simple Gödel encoding: assign to each symbol of the formula its ASCII code, assign to a formula the concatenation of the ASCII codes of its symbols.

# Gödel encodings

Example (Wikipedia): the formula

$$x = y \rightarrow y = x$$

written in ASCII as

$$x=y \Rightarrow y=x$$

corresponds to the sequence

120-061-121-032-061-062-032-121-061-120

of ASCII codes, and so it is assigned the number

120061121032061062032121061120



# Gödel's Gödel encoding

Let  $p_n$  denote the  $n$ -th prime number.

Gödel's encoding assigns to each symbol  $\lambda$  a number  $g(\lambda)$ , and to a sequence  $\lambda_1 \cdots \lambda_n$  of symbols the number

$$2^{g(\lambda_1)} \cdot 3^{g(\lambda_2)} \cdot 5^{g(\lambda_3)} \cdot \dots \cdot p_n^{g(\lambda_n)}$$

# What are Gödel encodings good for?

A formula  $F(x)$  over  $S_A$  with a free variable  $x$  defines a **property of numbers**: the property satisfied exactly by the numbers  $n$  such that  $F(s^n(0))$  is true in the canonical structure.

We can easily construct formulas  $Even(x)$ ,  $Prime(x)$ ,  $Power\_of\_two(x)$  ...

Via the encoding formulas “are” numbers, and so a formula also defines a property of formulas!

**numbers**  $\rightarrow$  **formulas**  
**formula**  $F(x)$   $\rightarrow$  **set of numbers**  $\rightarrow$  **set of formulas**

# Going further . . .

We can (less easily) construct formulas like

- *First\_symbol\_is\_* $\forall$ *(x)*
- *At\_least\_ten\_symbols*(x)
- *Closed*(x)
- . . .

that are true i.t.c.s. for  $x := s^n(0)$  iff the number  $n$  encodes a formula and the formula satisfies the corresponding property.

# And even further . . .

We can construct (even less easily) a formula

- $In_Q(x)$

that is true i.t.c.s. for  $x = s^n(0)$  iff the number  $n$  encodes a closed formula  $F$  such that  $F \in Q$ .

The reason is

$$F \in Q \text{ iff } Q1, \dots, Q9 \models F \text{ iff } Q1, \dots, Q9 \vdash F$$

and the derivation procedure amounts to symbol manipulation.

Same for any other set  $\mathcal{X}$  of axioms.

# Diagonal Lemma

Recall our goal: Given a set of axioms  $\mathcal{X} \subseteq \text{Arith}$ , construct a formula  $F \in \text{Arith}$  such that  $F \notin T_{\mathcal{X}}$

Let  $\underline{F}$  denote the term  $s^n(0)$  where  $n$  is the Gödel encoding of the formula  $F$ .

**Intuition:**  $\underline{F}$  is a “name” we give to  $F$

**Lemma (Diagonal Lemma):** Let  $\mathcal{X}$  be any set of axioms containing Q1, ... Q9. For every formula  $B(x)$  there is a closed formula  $G$  such that  $G \leftrightarrow B(\underline{G}) \in T_{\mathcal{X}}$ .

We call  $G$  the **Gödel formula** of  $B(x)$ .

We have:  $G$  true i.t.c.s if and only if  $G$  has property  $B$

**Intuition:**  $G$  asserts that  $G$  has property  $B$  (true or false i.t.c.s.!).

# Reaching the goal

**Theorem:** Let  $\mathcal{X}$  be any set of axioms containing Q1, ... Q9.  
Let  $G_{\mathcal{X}}$  be the Gödel formula of  $\neg In_{T_{\mathcal{X}}}(x)$ . Then  $G_{\mathcal{X}} \in Arith \setminus T_{\mathcal{X}}$ .

**Proof idea:** By definition,  $G_{\mathcal{X}}$  is true i.t.c.s iff  $G_{\mathcal{X}} \notin T_{\mathcal{X}}$ .

If  $G_{\mathcal{X}}$  is false i.t.c.s. then  $G_{\mathcal{X}} \in T_{\mathcal{X}}$ .

Since  $\mathcal{X} \subseteq Arith$ , we have  $G_{\mathcal{X}} \in Arith$ .

But then, by definition of Arith,  $G_{\mathcal{X}}$  is true i.t.c.s.

**Contradiction!**

So  $G_{\mathcal{X}}$  is true i.t.c.s., i.e.,  $G_{\mathcal{X}} \in Arith$ .

But then  $G_{\mathcal{X}} \notin T_{\mathcal{X}}$ , and so  $G_{\mathcal{X}} \in Arith \setminus T_{\mathcal{X}}$ . Done!

# Gödel's second incompleteness theorem

For any set of axioms  $\mathcal{X}$  containing Q1 we have  $0 = s(0) \notin T_{\mathcal{X}}$ , and so  $T_{\mathcal{X}}$  is consistent iff  $0 = s(0) \notin T_{\mathcal{X}}$ .

The **consistency formula** for  $\mathcal{X}$  is the formula  $\neg \text{In}_{T_{\mathcal{X}}}(0=s(0))$

**Intuition:** The consistency formula for  $\mathcal{X}$  states that  $T_{\mathcal{X}}$  is consistent.

**Theorem (Gödel's second incompleteness theorem):** Let  $\mathcal{X} \subseteq \text{Arith}$  be any set of axioms containing P. Then the consistency formula for  $\mathcal{X}$  does not belong to  $T_{\mathcal{X}}$ .

**Intuition:** the consistency of a theory cannot be derived from the axioms of the theory.

# Proving the Diagonal Lemma: Diagonalization

Let  $F(x)$  be a formula with a free variable  $x$ .  
The **diagonalization of  $F$**  is the closed formula

$$DiagF := \exists x x = \underline{F} \wedge F(x)$$

**Intuition:**  $DiagF$  asserts that  $F$  has property  $F$

Observe:  $DiagF$  and  $F(\underline{F})$  are logically equivalent, but they have different Gödel numbers.



# The representation theorem

**Theorem** : There is a formula  $Diag(x, y)$  such that for every formula  $F$

$$\forall y \quad Diag(\underline{F}, y) \leftrightarrow y = \underline{DiagF}$$

can be derived in Q (and so in P).

**Proof**: Omitted.

Observe: the theorem does not hold for every set of axioms. For instance, it does not hold for the system Q1-Q4, since in that system we cannot infer anything about the product function.

# Proof of the Diagonal Lemma I

**Lemma:** Let  $\mathcal{X}$  be any set of axioms containing Q1, ... Q9.

For every formula  $B(y)$  there is a closed formula  $G$  such that  $G \leftrightarrow B(\underline{g}) \in T_{\mathcal{X}}$ .

**Proof:** Let  $A(x) := \exists y (Diag(x, y) \wedge B(y))$  and let  $G := DiagA$ .

**Intuition:**  $G$  asserts that the diagonalization of  $A$  (the formula asserting that  $A$  satisfies  $A$ ) satisfies  $B$ .

Explicitly:

$$G := \exists x (x = \underline{A} \wedge A(x)) := \exists x (x = \underline{A} \wedge \exists y (Diag(x, y) \wedge B(y)))$$

# Proof of the Diagonal Lemma II

The formula  $G \leftrightarrow \exists y (Diag(\underline{A}, y) \wedge B(y))$  is valid, and so, since valid formulas belong to every theory, we have

$$G \leftrightarrow \exists y (Diag(\underline{A}, y) \wedge B(y)) \in T_{\mathcal{X}}$$

Since  $G := DiagA$ , we have by the representation theorem:

$$\forall y (Diag(\underline{A}, y) \leftrightarrow y = \underline{G}) \in T_{\mathcal{X}}$$

And so, since  $T_{\mathcal{X}}$  is closed under consequence, we get

$$G \leftrightarrow \exists y (y = \underline{G} \wedge B(y)) \in T_{\mathcal{X}}$$

and for the same reason

$$G \leftrightarrow B(\underline{G}) \in T_{\mathcal{X}}$$